

POSET LIMITS CAN BE TOTALLY ORDERED

JAN HLADKÝ, ANDRÁS MÁTHÉ, VIRESH PATEL, AND OLEG PIKHURKO

ABSTRACT. S. Janson [*Poset limits and exchangeable random posets*, Combinatorica **31** (2011), 529–563] defined limits of finite posets in parallel to the emerging theory of limits of dense graphs.

We prove that each poset limit can be represented as a kernel on the unit interval with the standard order, thus answering an open question of Janson. We provide two proofs: real-analytic and combinatorial. The combinatorial proof is based on a Szemerédi-type Regularity Lemma for posets which may be of independent interest.

1. INTRODUCTION

Given a class \mathcal{C} of finite structures and some measure $t(F, G)$ for $F, G \in \mathcal{C}$ of how frequently F appears in G as a substructure, one can say that a sequence $\{G_n\}_{n \in \mathbb{N}}$ converges if $\{t(F, G_n)\}_{n \in \mathbb{N}}$ converges for every $F \in \mathcal{C}$.

For example, if \mathcal{C} consists of finite graphs and t denotes the subgraph density, then we obtain the convergence of (dense) graphs whose systematic study was initiated by Lovász and Szegedy [13] and Borgs et al [4]. In particular, Lovász and Szegedy [13] showed that for every convergent sequence $\{G_n\}_{n \in \mathbb{N}}$ of graphs there is a measurable function $W : [0, 1]^2 \rightarrow [0, 1]$ (called *graphon*) such that for every graph F the limit of $t(F, G_n)$ as $n \rightarrow \infty$ is a certain integral involving W . In fact, many other parameters of G_n can be well approximated as $n \rightarrow \infty$ if we know W . This opens a general way of bringing analytic methods into the study of large finite graphs. Connections to other areas are established by alternative representations of “graph limits”: by reflection positive graph parameters (Lovász and Szegedy [13]), by positive flag algebra homomorphisms (Razborov [17]), and by partially exchangeable random arrays (Diaconis and Janson [6]).

The theory of graph limits has received a great deal of attention and has been extended to other structures as well, such as hypergraphs (Elek and Szegedy [8], see also Tao [21]

2010 *Mathematics Subject Classification.* 06A06, 28Axx.

Key words and phrases. Homomorphism density, ordered probability space, partially ordered set, poset kernel, Regularity Lemma.

J.H. was supported by an EPSRC fellowship.

A.M. was supported by EPSRC grant EP/G050678/1.

O.P. was supported by the European Research Council (grant agreement no. 306493) and the National Science Foundation of the USA (grant DMS-1100215).

and Austin [1]), permutations (Hoppen et al [9, 10]), functions on compact Abelian groups (Szegedy [19]), and other.

An analogous theory for limits of *posets* (i.e. partially ordered sets) was initiated by Brightwell and Georgiou [5] and further developed by Janson [11]. Let us state some of these results.

We represent a poset as a pair (P, \prec) where P is a finite *ground set* and \prec is a *strict order relation* (i.e. it is transitive and no $a, b \in P$ satisfy $a \prec b$ and $b \prec a$ simultaneously).

A map $f : P \rightarrow Q$ (not necessarily injective) is a *homomorphism* from (P, \prec) to (Q, \ll) if we have $f(x) \ll f(y)$ for every $x, y \in P$ with $x \prec y$. The *density* $t((P, \prec), (Q, \ll))$ is the number of homomorphisms from (P, \prec) to (Q, \ll) divided by the total number of possible maps $P \rightarrow Q$. In other words, it is the probability that a random map $P \rightarrow Q$ between the ground sets preserves the order relation.

Definition 1.1. A sequence of posets $\{(P_n, \prec_n)\}_{n \in \mathbb{N}}$ converges if $|P_n| \rightarrow \infty$ and

$$(1) \quad \{t((P, \prec), (P_n, \prec_n))\}_{n \in \mathbb{N}} \text{ converges for every poset } (P, \prec).$$

Remark 1.2. It is not hard to show (cf [13, Section 2.4]) that Definition 1.1 does not change if we modify t to be the density of induced and/or injective homomorphisms.

The potential usefulness of (1) comes from the result of Janson [11, Theorem 1.7] that for each convergent sequence there is an analytic limit object as follows. (See Section 2 for an overview of the measure theory notation that we use.)

Definition 1.3. An *ordered probability space* $(S, \mathcal{F}, \mu, \triangleleft)$ is a probability space (S, \mathcal{F}, μ) equipped with a strict order relation \triangleleft such that $\{(x, y) : x \triangleleft y\}$ is an $\mathcal{F} \otimes \mathcal{F}$ -measurable subset of $S \times S$.

Definition 1.4. A *(poset) kernel* is a 5-tuple $(S, \mathcal{F}, \mu, \triangleleft, W)$, where $(S, \mathcal{F}, \mu, \triangleleft)$ is an ordered probability space and W is an $\mathcal{F} \otimes \mathcal{F}$ -measurable function $S \times S \rightarrow [0, 1]$ such that, for all $x, y, z \in S$,

$$(2) \quad W(x, y) > 0 \Rightarrow x \triangleleft y,$$

$$(3) \quad W(x, y) > 0 \text{ and } W(y, z) > 0 \Rightarrow W(x, z) = 1.$$

In particular, it follows from Definition 1.4 that $W(x, y)W(y, x) = 0$ for every $x, y \in S$.

When no confusion arises, we may abbreviate (P, \prec) to P and $(S, \mathcal{F}, \mu, \triangleleft, W)$ to W . Also, we will usually say “kernel” instead of “poset kernel”.

Definition 1.5. The *density* of a poset (P, \prec) in a kernel $(S, \mathcal{F}, \mu, \triangleleft, W)$ is

$$(4) \quad t(P, W) := \int_{S^{|P|}} \prod_{\substack{a, b \in P \\ a \prec b}} W(x_a, x_b) \prod_{a \in P} d\mu(x_a).$$

There is some analogy between $t(P, Q)$ and $t(P, W)$. Namely, one can interpret the expression in the right-hand side of (4) as follows. Select random elements x_a of (S, \mathcal{F}, μ) indexed by P and let $x_a \ll x_b$ with probability $W(x_a, x_b)$, with all choices being mutually independent. Then $t(P, W)$ is exactly the probability that $x_a \ll x_b$ for all $a \prec b$ in P . In fact, the connection is much deeper as the following result shows.

Theorem 1.6 (Janson [11, Theorems 1.7 and 1.9(ii)]). *For every convergent sequence $\{P_n\}_{n \in \mathbb{N}}$ of posets there is a kernel $(S, \mathcal{F}, \mu, \triangleleft, U)$ such that*

$$(5) \quad t(P, U) = \lim_{n \rightarrow \infty} t(P, P_n), \quad \text{for every poset } P.$$

Moreover, we can assume in (5) that

$$(6) \quad (S, \mathcal{F}, \mu) = ([0, 1], \mathcal{B}, \lambda)$$

is the unit interval with the Lebesgue measure λ on the Borel σ -algebra \mathcal{B} .

Also, the converse of Theorem 1.6 was established in [11]: for every kernel U there is a sequence of posets $\{P_n\}_{n \in \mathbb{N}}$ that satisfies (5). In fact, the sampling procedure informally described after (4) yields such a sequence with probability 1.

Although we can require that (6) holds, the proof in [11] gives no control over the order relation \triangleleft . This prompted Janson [11, Problem 1.10] to ask if one can always take $([0, 1], \mathcal{B}, \lambda, <)$ with the standard order $<$ in Theorem 1.6. In a later paper [12], Janson answered his question for convergent sequences of interval orders (see also [11, Theorem 1.9(iii)] for a related result). Here we give the affirmative answer in the general case.

Theorem 1.7. *For every convergent sequence $\{P_n\}_{n \in \mathbb{N}}$ of posets there is a kernel $([0, 1], \mathcal{B}, \lambda, <, U)$ such that (5) holds.*

In fact, we provide two different proofs of Theorem 1.7.

One goes via a Regularity Lemma for posets that we prove in Section 6. Our lemma finds a partition $P = V_1 \cup \dots \cup V_k$ which is ε -regular with respect to the underlying graph of (P, \prec) and has the additional property that all \prec -relations between parts go “forward” only. Having such a Regularity Lemma, we follow the method of Lovász and Szegedy [13, 14] to construct a kernel U by taking the “limit” of ε -regular partitions as $\varepsilon \rightarrow \infty$. The above “forward” property allows us to ensure that $U(x, y) = 0$ whenever $x \geq y$, thus proving Theorem 1.7. We expect our Regularity Lemma to have further applications.

The other proof of Theorem 1.7 is real-analytic. Actually, we prove a somewhat stronger result (Theorem 1.9 below). In order to state it, we have to give some further definitions.

Let $(S, \mathcal{F}, \mu, \triangleleft, W)$ be a kernel. We call it *strict* if $W(x, y) > 0$ for every $x, y \in S$ with $x \triangleleft y$. (Thus a kernel is strict if the two sides of (2) are equivalent.) Kernel axioms imply that if define $\triangleleft' := \{(x, y) \in S^2 : W(x, y) > 0\}$ then $(S, \mathcal{F}, \mu, \triangleleft')$ is an ordered probability space on which W is a strict kernel. Clearly, this change does not affect (4). Thus, we can additionally assume in Theorem 1.6 that U is strict, see [11, Remark 1.2].

Definition 1.8. An *inclusion* between ordered probability spaces $(S, \mathcal{F}, \mu, \triangleleft)$ and $(S', \mathcal{F}', \mu', \triangleleft')$ is a measure preserving function $f : S \rightarrow S'$ such that the set

$$(7) \quad \{ (x, y) \in S^2 : x \triangleleft y, f(x) \not\triangleleft' f(y) \}$$

has $\mu \otimes \mu$ -measure zero. Additionally, if we have a kernel U on $(S', \mathcal{F}', \mu', \triangleleft')$ then its *pull-back along f* is the function $U^f : S^2 \rightarrow [0, 1]$, defined by $U^f(x, y) := U(f(x), f(y))$ for $x, y \in S$.

Theorem 1.9. *For every strict kernel $(S, \mathcal{F}, \mu, \triangleleft, W)$ such that (S, \mathcal{F}, μ) is atomless, there is a kernel $([0, 1], \mathcal{B}, \lambda, <, U)$ and an inclusion $f : (S, \mathcal{F}, \mu, \triangleleft) \rightarrow ([0, 1], \mathcal{B}, \lambda, <)$ such that W is equal to the pull-back U^f almost everywhere.*

Since f in Theorem 1.9 is measure preserving, we necessarily have that $t(P, U) = t(P, W)$ for every poset P , that is, U and W represent the same poset limit. Thus Theorem 1.7 follows from Theorems 1.6 and 1.9.

The notion of a pull-back plays an important role in the theory of graphons. Hopefully, our Theorem 1.9 will be generally useful when studying poset kernels. For example, if the studied kernel property behaves well with respect to taking pull-backs, then one can operate with the function U that satisfies Theorem 1.9 instead of the 5-tuple $(S, \mathcal{F}, \mu, \triangleleft, W)$.

Given an ordered probability space $(S, \mathcal{F}, \mu, \triangleleft)$, the indicator function I_{\triangleleft} of the order relation \triangleleft is clearly a strict kernel on it. Thus Theorem 1.9 has the following direct corollary.

Theorem 1.10. *Every atomless ordered probability space $(S, \mathcal{F}, \mu, \triangleleft)$ can be included into $([0, 1], \mathcal{B}, \lambda, <)$.* □

Theorem 1.10 can be viewed as a measure theoretic analogue of the statement that every poset can be totally ordered. While extending this to infinite partially ordered sets is an easy application of Zorn's lemma, the main content of Theorem 1.10 is that this total ordering can be done in a “measurable” way. Interestingly, the limit of totally ordered increasing posets happens to be our universal target space $([0, 1], \mathcal{B}, \lambda, <)$ with the indicator function $I_{<}$ as its kernel.

This paper is organised as follows. Section 2 describes the measure theory notation that we frequently use. Section 3 presents some auxiliary analytic lemmas, thus making the flow of argument in the later sections smoother. Although Theorem 1.10 is a direct consequence of Theorem 1.9, we prove it first in Section 4. Then, in Section 5, we show how Theorem 4.1 (a version of Theorem 1.10) implies Theorem 1.9. Our Regularity Lemma for posets is stated and proved in (combinatorial) Section 6 which can be read independently of the other sections. We show how this Regularity Lemma gives an alternative proof of Theorem 1.7 in Section 7. Finally, Section 8 contains some concluding remarks, including examples that certain strengthenings of our results are not possible.

2. MEASURE THEORY NOTATION

Let us give some notation that we are going to use frequently. We do not define many standard concepts of measure theory but refer the reader to Bogachev's book [2] whose notation we generally follow. We try to provide sufficient references and explanations so that this paper is accessible to combinatorialists who do not have a strong background in measure theory.

Let $\mathbb{N} = \{1, 2, \dots\}$ and \mathbb{R} be the sets of respectively natural and real numbers. When we consider a subset of \mathbb{R} , typically the unit interval $[0, 1]$, we will denote the σ -algebra of its Borel subsets by \mathcal{B} and the Lebesgue measure by λ . For a family \mathcal{X} of sets, let $\sigma(\mathcal{X})$ denote the σ -algebra generated by \mathcal{X} . Let I_X denote the indicator function of a set X (that is, $I_X(x)$ is 1 if $x \in X$ and 0 otherwise).

Let (S, \mathcal{F}, μ) be a probability space. As it is standard in measure theory, a real-valued function f on S is called \mathcal{F} -measurable if it is $(\mathcal{F}, \mathcal{B})$ -measurable. We denote by \mathcal{F}_μ the completion of \mathcal{F} with respect to the measure μ .

We say that a property holds (\mathcal{F}, μ) -almost everywhere (and abbreviate this to (\mathcal{F}, μ) -a.e.) if the set of points of S where it fails belongs to \mathcal{F}_μ and has μ -measure zero. When the underlying measure space is understood, we just write "a.e." In some rare cases when we consider more than one σ -algebra on the same set, the bare term "a.e." refers to the largest σ -algebra.

We call two sets or two functions *equivalent* (and use the symbol \sim) if they coincide a.e. The *Fréchet-Nikodym distance* between two sets $A, B \in \mathcal{F}_\mu$ is $\mu(A \Delta B)$; it is in general a pseudo-metric (it satisfies the Triangle Inequality but may evaluate to 0 for $A \neq B$). The space (S, \mathcal{F}, μ) is called *separable* if \mathcal{F} has a countable subset which is dense with respect to the Fréchet-Nikodym distance.

Let $\mathcal{A} \subseteq \mathcal{F}$ be another σ -algebra and let f be an integrable real-valued function on (S, \mathcal{F}, μ) . The *conditional expectation* $\mathbb{E}(f|\mathcal{A})$ is the set of all \mathcal{A} -measurable functions $g : S \rightarrow \mathbb{R}$ such that for every bounded \mathcal{A} -measurable function $h : S \rightarrow \mathbb{R}$ we have

$$(8) \quad \int h(x)g(x) \, d\mu(x) = \int h(x)f(x) \, d\mu(x).$$

As it is well-known, every two functions in $\mathbb{E}(f|\mathcal{A})$ are equivalent and it is enough to check (8) for $\{0, 1\}$ -valued h only, i.e. that

$$(9) \quad \int_A g(x) \, d\mu(x) = \int_A f(x) \, d\mu(x) \quad \text{for all } A \in \mathcal{A}.$$

We refer the reader to [2, Section 10.1.1] for some basic properties of conditional expectation. We may treat $\mathbb{E}(f|\mathcal{A})$ as a single function (rather as a set of functions), when the studied property does not depend on the choice of a representative.

Let (S', \mathcal{F}', μ') be another probability space. A map $f : S \rightarrow S'$ is *measure preserving* if f is $(\mathcal{F}, \mathcal{F}')$ -measurable and, for every $A \in \mathcal{F}'$, we have $\mu(f^{-1}(A)) = \mu'(A)$. The products of σ -algebras and measures are denoted by $\mathcal{F} \otimes \mathcal{F}'$ and $\mu \otimes \mu'$. We use the shorthand $\mathcal{F} \bar{\otimes} \mathcal{F}'$ for $(\mathcal{F} \otimes \mathcal{F}')_{\mu \otimes \mu'}$, the completion of $\mathcal{F} \otimes \mathcal{F}'$ with respect to $\mu \otimes \mu'$. We will be using Fubini's theorem ([2, Theorem 3.4.4]) very frequently, often without explicitly mentioning it. Let us stress that one has to be careful when dealing with products of σ -algebras and measures. For example, the product of two complete measure spaces is not complete in general. Also, see Exercises 44–45, 49–51, and 55 in [2, Section 3.10] for counterexamples to some “plausible” statements related to Fubini's theorem.

Let us give some kernel-specific definitions (when the underlying ordered probability space $(S, \mathcal{F}, \mu, \triangleleft)$ is understood). For $A \subseteq S$, let $A^c := S \setminus A$ denote the complement of A . For $X \in \mathcal{F} \otimes \mathcal{F}$, we define

$$(10) \quad \mu_{\triangleleft}(X) := \int_X I_{\triangleleft}(a, b) \, d\mu(a) \, d\mu(b).$$

For a 2-variable function $W : S^2 \rightarrow \mathbb{R}$ and $y \in S$, the *slice function* $W_y : S \rightarrow \mathbb{R}$ is defined by $W_y(x) := W(x, y)$. We call $W : S \times S \rightarrow [0, 1]$ an *almost (poset) kernel* if W is $\mathcal{F} \bar{\otimes} \mathcal{F}$ -measurable and the kernel axioms (2) and (3) hold for a.e. triple (x, y, z) in (S, \mathcal{F}, μ) ³.

Although our Theorem 1.9 takes a kernel W as input and then produces another kernel U , we have to deal with almost kernels at intermediate stages of the proof. (For example, the pull-back U^f in Theorem 1.9 is generally an almost kernel.)

3. AUXILIARY ANALYTIC LEMMAS

Here we present some auxiliary results that we will need later.

Janson [11] proved that one can transform an almost kernel $(S, \mathcal{B}, \lambda, \triangleleft, W)$ with $S \subseteq \mathbb{R}$ into a kernel $(S, \mathcal{B}, \lambda, \triangleleft', W')$ with $W' \sim W$ and some \triangleleft' . We show that in the special case of the unit interval with the standard order, one can also keep the order relation intact.

Lemma 3.1. *Let $([0, 1], \mathcal{B}, \lambda, <, U)$ be an almost kernel. Then there is $U' \sim U$ such that $([0, 1], \mathcal{B}, \lambda, <, U')$ is a kernel.*

Proof. First, we choose a $\mathcal{B} \otimes \mathcal{B}$ -measurable function $U_0 \sim U$; it exists by [2, Proposition 2.1.11]. Then we proceed in a similar fashion as is done by Janson [11, Pages 547–548], so we will be rather brief. We refer the reader to [2, Section 5] for the definitions and basic properties of Lebesgue and density points.

We define $U_1 : [0, 1]^2 \rightarrow [0, 1]$ by $U_1(x, y) := U_0(x, y)$ if $(x, y) \in [0, 1]^2$ is a Lebesgue point of U_0 . Next, if (x, y) is a density point of the set $\{(x, y) : U_0(x, y) = 1\}$, then let $U_1(x, y) := 1$. (Recall that a density point need not belong to the set itself.) For all other pairs $(x, y) \in [0, 1]^2$, we define $U_1(x, y) := 0$. Note that $U_1 \sim U_0$ and U_1 is still Borel.

We claim that U_1 is a kernel on $([0, 1], \mathcal{B}, \lambda, <)$. Suppose that $U_1(x, y) > 0$. Then for every sufficiently small $\varepsilon > 0$, we have $U_0(x', y') > 0$ for most points $(x', y') \in (x \pm \varepsilon) \times (y \pm \varepsilon)$, where $(x \pm \varepsilon)$ denotes the intersection of the open interval $(x - \varepsilon, x + \varepsilon)$ with $[0, 1]$. In particular, $x' < y'$ for most of these pairs and therefore $x < y$.

Now suppose that $U_1(x, y) > 0$ and $U_1(y, z) > 0$. Then for every sufficiently small $\varepsilon > 0$, we have $U_0(x', y') > 0$ for most points $(x', y') \in (x \pm \varepsilon) \times (y \pm \varepsilon)$ and $U_0(y', z') > 0$ for most points $(y', z') \in (y \pm \varepsilon) \times (z \pm \varepsilon)$. This implies that we have $U_0(x', z') = 1$ for most points $(x', z') \in (x \pm \varepsilon) \times (z \pm \varepsilon)$. Thus (x, z) is a density point of $\{(x, y) : U_0(x, y) = 1\}$ and therefore $U_1(x, z) = 1$, as required. \square

Lemma 3.2 (Borgs, Chayes, and Lovász [3, Lemma 3.4]). *Let (S, \mathcal{F}) and (S', \mathcal{F}') be measurable spaces, and let $W : S \times S' \rightarrow \mathbb{R}$ be a bounded $\mathcal{F} \otimes \mathcal{F}'$ -measurable function. Then there exist countably generated σ -algebras $\mathcal{F}_0 \subseteq \mathcal{F}$ and $\mathcal{F}'_0 \subseteq \mathcal{F}'$ such that W is $\mathcal{F}_0 \otimes \mathcal{F}'_0$ -measurable.* \square

Lemma 3.3. *Let $f : S \rightarrow S'$ be an inclusion of ordered probability spaces $(S, \mathcal{F}, \mu, \triangleleft)$ and $(S', \mathcal{F}', \mu', \triangleleft')$. Let W be a kernel on S such that $W \sim \mathbb{E}(W|\mathcal{A} \otimes \mathcal{A})$, where $\mathcal{A} := f^{-1}(\mathcal{F}')$. Then there is an almost kernel U on $(S', \mathcal{F}', \mu', \triangleleft')$ with $W \sim U^f$.*

Proof. We construct U following the argument of Borgs, Chayes, and Lovász [3, Lemma 3.1] rather closely.

Define a function ν from $\mathcal{F}' \times \mathcal{F}' := \{A \times B \mid A, B \in \mathcal{F}'\}$ to \mathbb{R} by

$$\nu(A \times B) := \int_{f^{-1}(A) \times f^{-1}(B)} W(x, y) d\mu(x) d\mu(y), \quad A, B \in \mathcal{F}'.$$

This is a countably additive function on the semialgebra $\mathcal{F}' \times \mathcal{F}'$ which uniquely extends to a measure ν on $\mathcal{F}' \otimes \mathcal{F}'$ by [2, Proposition 1.3.10]. It is easy to see that this measure ν is absolutely continuous with respect to $\mu' \otimes \mu'$. Hence, the Radon-Nikodym derivative

$$U := \frac{d\nu}{d(\mu' \otimes \mu')}$$

exists ([2, Theorem 3.2.2]). Namely, U is a $\mu' \otimes \mu'$ -integrable function such that for every $X \in \mathcal{F}' \otimes \mathcal{F}'$ we have $\nu(X) = \int_X U d(\mu' \otimes \mu')$.

The last identity implies (given that f is measure preserving and that $0 \leq W \leq 1$) that the set

$$\{(x, y) \in S' \times S' : U(x, y) > 1 \text{ or } U(x, y) < 0\}$$

has measure zero. By changing U on a null set, we can assume that U is $\mathcal{F}' \otimes \mathcal{F}'$ -measurable (see [2, Proposition 2.1.11]) and that the values of U belong to $[0, 1]$. In particular, the pull-back U^f is $\mathcal{A} \otimes \mathcal{A}$ -measurable. Moreover, for any $Y \in \mathcal{A} \otimes \mathcal{A}$, say $Y = (f \times f)^{-1}(X)$, we have that

$$(11) \quad \int_Y U^f d(\mu \otimes \mu) = \int_X U d(\mu' \otimes \mu') = \int_Y W d(\mu \otimes \mu).$$

By (9), we conclude that $U^f \in \mathbb{E}(W|\mathcal{A} \otimes \mathcal{A})$. Thus U^f is a.e. equal to W by the assumption of the lemma.

Let us verify that U is an almost kernel. First, consider the set

$$X := \{(x, y) \in S' \times S' : x \not\preccurlyeq y, U(x, y) > 0\} \in \mathcal{F}' \otimes \mathcal{F}'$$

of points where the first kernel axiom (2) fails for U . By (11), the integral of U over X is the same as the integral of W over $Y := (f \times f)^{-1}(X)$. Since f is an inclusion, we have $\mu_{\triangleleft}(Y) = 0$, where μ_{\triangleleft} is defined by (10). Since W is a kernel, it is zero a.e. on Y . It follows that X has measure zero, that is, U satisfies (2) a.e.

Define $u(x, y, z) := U(x, y)U(y, z)(1 - U(x, z))$. Since $U^f \sim W$ and W is a kernel, we have $u^f \sim 0$. Since f is measure-preserving, we have $\int u = \int u^f = 0$. The non-negativity of u implies that $u \sim 0$, that is, U satisfies (3) a.e. \square

Remark 3.4. The conditional expectation of a kernel need not be an almost kernel. For example, let $S := \{a, b, c, d\}$ with $\mathcal{F} := 2^S$ and μ being the uniform measure. Let $\mathcal{A} \subseteq \mathcal{F}$ be obtained by “gluing” b and c together. Let $a \triangleleft b$ and $c \triangleleft d$ be all order relations and let $W := I_{\triangleleft}$. Then any $U \in \mathbb{E}(W|\mathcal{A} \otimes \mathcal{A})$ satisfies $U(a, b) = 1/2$, $U(b, d) = 1/2$, and $U(a, d) = 0$ and cannot be an almost kernel. Also, pull-backs do not preserve (almost) kernels in general: for example, consider the pull-back of I_{\triangleleft} with respect to the identity inclusion of $([0, 1], \mathcal{B}, \lambda, \emptyset)$ into $([0, 1], \mathcal{B}, \lambda, \triangleleft)$.

Lemma 3.5. *Let (S, \mathcal{F}, μ) be a probability space. Let $\mathcal{A} \subseteq \mathcal{F}$ be another σ -algebra such that (S, \mathcal{A}, μ) is separable. Let $W : S^2 \rightarrow \mathbb{R}$ be a bounded $\mathcal{F} \otimes \mathcal{F}$ -measurable function. Let $g \in \mathbb{E}(W|\mathcal{A} \otimes \mathcal{F})$. Then, for a.e. $y \in S$, we have that $g_y \sim \mathbb{E}(W_y|\mathcal{A})$.*

Proof. By definition, g is $\mathcal{A} \otimes \mathcal{F}$ -measurable. It follows by [2, Proposition 3.3.2] that the slice function g_y is \mathcal{A} -measurable for every $y \in S$.

Fix $A \in \mathcal{A}$. By the definition of conditional expectation, we have that $\int_{A \times B} g = \int_{A \times B} W$ for every $B \in \mathcal{F}$. Likewise,

$$(12) \quad \int_A \mathbb{E}(W_y|\mathcal{A}) = \int_A W_y, \quad \text{for every } y \in S.$$

By Fubini’s theorem, the latter function is integrable as a function of y . Moreover,

$$\int_B \left(\int_A W_y(x) d\mu(x) \right) d\mu(y) = \int_{A \times B} W = \int_{A \times B} g = \int_B \left(\int_A g_y(x) d\mu(x) \right) d\mu(y).$$

Since $B \in \mathcal{F}$ was arbitrary, [2, Corollary 2.5.4] gives that $\int_A W_y = \int_A g_y$ for a.e. y . Let us remove all exceptional points y when A runs over a dense countable subset $\{A_1, A_2, \dots\} \subseteq \mathcal{A}$ in (S, \mathcal{A}, μ) as well as those y for which $\|g_y\|_\infty > \|W\|_\infty$ or $\|W_y\|_\infty > \|W\|_\infty$. It is easy to see that the remaining set Y has measure 1.

Fix any $y \in Y$. For every $A \in \mathcal{A}$ we have that

$$\left| \int_A W_y - \int_A g_y \right| \leq \left| \int_A W_y - \int_{A_i} W_y \right| + \left| \int_{A_i} W_y - \int_{A_i} g_y \right| \leq 4 \|W\|_\infty \mu(A \triangle A_i).$$

Since the right-hand side can be made arbitrarily small by choosing a suitable A_i , we conclude that $\int_A W_y = \int_A g_y$. Since $A \in \mathcal{A}$ was arbitrary and both $\mathbb{E}(W_y | \mathcal{A})$ and g_y are \mathcal{A} -measurable, they coincide a.e. by (12). The lemma is proved as $\mu(S \setminus Y) = 0$. \square

4. PROOF OF THEOREM 1.10

Let $(S, \mathcal{F}, \mu, \triangleleft)$ be given. Lemma 3.2, when applied to the indicator function I_{\triangleleft} , returns two countably generated σ -algebras $\mathcal{F}_0, \mathcal{F}'_0 \subseteq \mathcal{F}$. Let $\mathcal{F}' := \sigma(\mathcal{F}_0 \cup \mathcal{F}'_0)$ be the σ -algebra on S generated by $\mathcal{F}_0 \cup \mathcal{F}'_0$. By enlarging a set of generators of \mathcal{F}' by adding a countably many elements of \mathcal{F} , we can additionally make (S, \mathcal{F}', μ) atomless.

Clearly, if we prove Theorem 1.10 for this new space $(S, \mathcal{F}', \mu, \triangleleft)$, then the same inclusion f will work for the original one (as $\mathcal{F}' \subseteq \mathcal{F}$). Thus, without loss of generality, let us assume that \mathcal{F} is countably generated. It easily follows (see e.g. Exercise 1.12.102 and its hint in [2]) that (S, \mathcal{F}, μ) is separable. Thus it is enough to prove the following theorem (whose last claim will be needed later in Section 5).

Theorem 4.1. *Let $(S, \mathcal{F}, \mu, \triangleleft)$ be an ordered probability space such that (S, \mathcal{F}, μ) is atomless and separable. Then there is an inclusion $f : (S, \mathcal{F}, \mu, \triangleleft) \rightarrow ([0, 1], \mathcal{B}, \lambda, <)$ such that every set $A \in \mathcal{F}$ with $\mu_{\triangleleft}(A \times A^c) = 0$ belongs to $(f^{-1}(\mathcal{B}))_{\mu}$, the completion of $f^{-1}(\mathcal{B})$ with respect to the measure μ .*

So we prove Theorem 4.1 now.

Claim 1. Let $B \in \mathcal{F}$ with $\mu(B) > 0$. Then there exists $A \in \mathcal{F}$ such that $\mu_{\triangleleft}(A \times A^c) = 0$, $\mu(B \cap A) > 0$, and $\mu(B \cap A^c) > 0$.

Proof of Claim. Let $\triangleright_x := \{y \in S : y \triangleright x\} \in \mathcal{F}$ be the *strict upper shadow* of $x \in S$ and let

$$B' := \{x \in B : \mu(B \cap \triangleright_x) > 0\}.$$

First, suppose that $\mu(B') > 0$. Clearly, $\mu_{\triangleleft}(B' \times B') \leq \mu(B')^2/2$. By Fubini's theorem, there is $x \in B'$ with $\mu(B' \cap \triangleright_x) \leq \mu(B')/2$. Clearly, $A := \triangleright_x$ has the required properties.

If $\mu(B') = 0$, then $\mu_{\triangleleft}(B \times B) = 0$ by Fubini's theorem. Since \mathcal{F} is atomless, it contains $A' \subseteq B$ with $0 < \mu(A' \cap B) < \mu(B)$. The function $a(x) := \mu(\{y \in A' : y \triangleleft x\})$ is \mathcal{F} -measurable by [2, Corollary 3.3.3]. The set $X := \{x \in S : a(x) > 0\} \in \mathcal{F}$ is clearly up-closed with respect to \triangleleft and it intersects B in a set of measure 0 by Fubini's theorem. It is easy to see that $A := A' \cup X$ satisfies the claim. \square

Let

$$\mathcal{T} := \{A \in \mathcal{F} : 0 < \mu(A) < 1, \mu_{\triangleleft}(A \times A^c) = 0\}.$$

By Claim 1, \mathcal{T} is non-empty and, moreover, infinite. Since (S, \mathcal{F}, μ) is separable, we can choose a countable subset $\{A_1, A_2, \dots\} \subseteq \mathcal{T}$ which is dense in \mathcal{T} with respect to the Fréchet-Nikodym distance.

We define f so that it satisfies the following properties:

$$\begin{aligned}
 (13) \quad f(A_1^c) &\subseteq [0, \mu(A_1^c)], \\
 f(A_1) &\subseteq [\mu(A_1^c), 1], \\
 f(A_1^c \cap A_2^c) &\subseteq [0, \mu(A_1^c \cap A_2^c)], \\
 f(A_1^c \cap A_2) &\subseteq [\mu(A_1^c \cap A_2^c), \mu(A_1^c)], \\
 f(A_1 \cap A_2^c) &\subseteq [\mu(A_1^c), \mu(A_1^c) + \mu(A_1 \cap A_2^c)], \\
 f(A_1 \cap A_2) &\subseteq [\mu(A_1^c) + \mu(A_1 \cap A_2^c), 1],
 \end{aligned}$$

and so on. Specifically, for a (finite or infinite) binary sequence $\mathbf{b} = (b_1, b_2, \dots)$, let

$$\begin{aligned}
 A_{\mathbf{b}} &:= \bigcap \{A_i^c : b_i = 0\} \cap \bigcap \{A_i : b_i = 1\}, \\
 S_{\mathbf{b}} &:= A_{\mathbf{b}} \cup \bigcup \{A_{b_1, \dots, b_{i-1}, 0} : b_i = 1\} = \bigcup \{A_{\mathbf{b}'} : \mathbf{b}' \leq_{\text{lex}} \mathbf{b}\},
 \end{aligned}$$

where \leq_{lex} denotes the lexicographical order (which we apply only to two binary sequences of the same length). Next, for $x \in S$ define $\mathbf{b}(x) := (b_1(x), b_2(x), \dots) \in \{0, 1\}^{\mathbb{N}}$ by $b_i = I_{A_i}$ for $i \in \mathbb{N}$. Thus $\mathbf{b}(x)$ is the unique infinite sequence with $x \in A_{\mathbf{b}(x)}$. Finally, we define

$$f(x) := \mu(S_{\mathbf{b}(x)}).$$

Claim 2. The function f is \mathcal{F} -measurable.

Proof of Claim. The function $(x, y) \mapsto I_{\leq_{\text{lex}}}(\mathbf{b}(x), \mathbf{b}(y))$ is $\mathcal{F} \otimes \mathcal{F}$ -measurable: the pre-image of 0 is

$$\bigcup_{i \in \mathbb{N}} \bigcup_{b_1, \dots, b_{i-1}} (A_{b_1, \dots, b_{i-1}, 1} \times A_{b_1, \dots, b_{i-1}, 0}) \in \mathcal{F} \otimes \mathcal{F}.$$

Thus $f(x) = \int I_{\text{lex}}(\mathbf{b}(y), \mathbf{b}(x)) d\mu(y)$ is \mathcal{F} -measurable by [2, Corollary 3.3.3]. \square

Claim 3. For every $a \in [0, 1]$ and every infinite \mathbf{b} , both sets $f^{-1}(a)$ and $A_{\mathbf{b}}$ belong to \mathcal{F} and have μ -measure zero.

Proof of Claim. We have $f^{-1}(a) \in \mathcal{F}$ by Claim 2 and $A_{\mathbf{b}} \in \mathcal{F}$ because each A_i is in \mathcal{F} . Each of $f^{-1}(a)$ and $A_{\mathbf{b}}$ is a null set because otherwise some A_i would cut it into two parts of positive measure, which is clearly impossible. \square

Claim 4. Let $|\mathbf{b}|$ denote the length of the sequence \mathbf{b} . Then

$$\lim_{n \rightarrow \infty} \sup_{|\mathbf{b}|=n} \mu(A_{\mathbf{b}}) = 0.$$

Proof of Claim. Assume to the contrary this \limsup is $\varepsilon > 0$. Then, by König's lemma, there exists an infinite sequence $\mathbf{b} = (b_1, b_2, \dots)$ such that $\mu(A_{b_1, \dots, b_n}) \geq \varepsilon$ for every n . (Note that $A_{b_1, \dots, b_n} \supseteq A_{b_1, \dots, b_{n+1}}$.) As $A_{\mathbf{b}} = \bigcap_{n=1}^{\infty} A_{b_1, \dots, b_n}$, we conclude that $\mu(A_{\mathbf{b}}) \geq \varepsilon > 0$, contradicting Claim 3. \square

Claim 5. The set $\{\mu(S_{\mathbf{b}}) : |\mathbf{b}| \text{ is finite}\}$ is dense in $[0, 1]$.

Proof of Claim. Consider the binary sequences of length n . Notice that, for any finite \mathbf{b} ,

$$\mu(S_{\mathbf{b}}) = \sum_{\mathbf{b}' \leq_{\text{lex}} \mathbf{b}} \mu(A_{\mathbf{b}'}),$$

$\mu(S_{1,1,\dots,1}) = 1$, and that $\mu(S_{0,0,\dots,0}) = \mu(A_{0,0,\dots,0})$ (which tends to 0 by Claim 3). Let $\mathbf{b}' \leq_{\text{lex}} \mathbf{b}''$ be two sequences of length n which are consecutive in \leq_{lex} . Then

$$\mu(S_{\mathbf{b}''}) - \mu(S_{\mathbf{b}'}) = \mu(A_{\mathbf{b}''}) \leq \sup_{|\mathbf{b}|=n} \mu(A_{\mathbf{b}}).$$

Combining this with Claim 4 gives the statement. \square

Claim 6. The function f is measure preserving.

Proof of Claim. Claim 5 implies that the intervals $[0, \mu(S_{\mathbf{b}})]$, where \mathbf{b} runs over finite binary sequences, generate the Borel σ -algebra. Thus is enough to show that for every finite \mathbf{b} we have $\mu(f^{-1}([0, a])) = a$, where $a := \mu(S_{\mathbf{b}})$. The latter identity follows from the fact that the symmetric difference of $S_{\mathbf{b}}$ and $f^{-1}([0, a])$ is a subset of $f^{-1}(a)$ and therefore has measure zero by Claim 3. \square

The set $Y := \{(x, y) \in S^2 : x \triangleleft y, f(x) > f(y)\}$ is a subset of $\bigcup_{i=1}^{\infty} (A_i \times A_i^c) \in \mathcal{F} \otimes \mathcal{F}$. But the latter set has $\mu \otimes \mu$ -measure zero by the definition of A_i 's. Thus Y also has measure zero. Next, consider the set $Y_0 := \{(x, y) \in S^2 : f(x) = f(y)\}$. Since f is \mathcal{F} -measurable, we have $Y_0 \in \mathcal{F} \otimes \mathcal{F}$. Every slice of Y_0 has measure zero by Claim 3. By Fubini's theorem, Y_0 has itself measure zero. We conclude that f is an inclusion.

Finally, take an arbitrary $A \in \mathcal{F}$ with $\mu_{\triangleleft}(A \times A^c) = 0$. For every $i \in \mathbb{N}$ there is a set $A_{n_i} \in \mathcal{T}$ such that $\mu(A_{n_i} \triangle A) < 2^{-i}$. Since $A_{n_i} \triangle f^{-1}(X) \subseteq f^{-1}(Y)$, where X is a finite union of intervals and Y is the set of their endpoints, we have by Claim 3 that A_{n_i} is $(f^{-1}(\mathcal{B}))_{\mu}$ -measurable. This implies that A is $(f^{-1}(\mathcal{B}))_{\mu}$ -measurable: indeed, for

$$(14) \quad A' := \limsup_{i \rightarrow \infty} A_{n_i} = \bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} A_{n_j} \in (f^{-1}(\mathcal{B}))_{\mu}$$

we have $\mu(A' \triangle A) = 0$. This finishes the proof of Theorem 4.1 (and Theorem 1.10).

5. PROOF OF THEOREM 1.9

As in Theorem 1.10, we can assume that \mathcal{F} is separable. Apply Theorem 4.1 to $(S, \mathcal{F}, \mu, \triangleleft)$ to obtain an inclusion $f : S \rightarrow [0, 1]$. As we will see later, the same f will work in Theorem 1.9. (Thus, rather interestingly, f can be chosen independently of W in Theorem 1.9 if \mathcal{F} is separable.) Let

$$\mathcal{A} := (f^{-1}(\mathcal{B}))_{\mu}.$$

Since f is \mathcal{F} -measurable, we have that $\mathcal{A} \subseteq \mathcal{F}_{\mu}$.

We would like to apply Lemma 3.3. In order to do so, we have to verify first that $W \sim E(W|\mathcal{A} \otimes \mathcal{A})$. (Note that $E(W|\mathcal{A} \otimes \mathcal{A}) \sim E(W|f^{-1}(\mathcal{B}) \otimes f^{-1}(\mathcal{B}))$.)

Claim 1. For every y , the slice function W_y is \mathcal{A} -measurable.

Proof of Claim. Fix any $a \in [0, 1]$. For every $y \in S$, the set

$$A := W_y^{-1}((a, 1]) = \{x \in S : W(x, y) > a\}$$

is in \mathcal{F} . Since W is a strict kernel, we have $\mu_{\triangleleft}(A^c \times A) = 0$ for every y . By the second part of Theorem 4.1, $A \in \mathcal{F}$ belongs in fact to \mathcal{A} . Since intervals $(a, 1]$ generate the Borel σ -algebra, the claim follows. \square

Thus W and $\mathbb{E}(W|\mathcal{A} \otimes \mathcal{F}_\mu)$ are both $\mathcal{F} \bar{\otimes} \mathcal{F}$ -measurable. (Note that $\mathcal{F} \bar{\otimes} \mathcal{F} = \mathcal{F}_\mu \bar{\otimes} \mathcal{F}_\mu$.) Also their y -slices are equivalent for a.e. y by Lemma 3.5. By Fubini's theorem, the subset of S^2 where these two functions differ has $\mu \otimes \mu$ -measure zero. In other words, W is $\mathcal{A} \bar{\otimes} \mathcal{F}$ -measurable and, by symmetry, $\mathcal{F} \bar{\otimes} \mathcal{A}$ -measurable.

Claim 2.

$$W \sim \mathbb{E}(W|\mathcal{A} \otimes \mathcal{A}).$$

Proof of Claim. We follow the argument of Borgs, Chayes, and Lovász [3, Section 3.3.5]. Let $\widetilde{W} \in \mathbb{E}(W|\mathcal{A} \otimes \mathcal{A})$. It is enough to prove that for every $A, B \in \mathcal{F}$,

$$\int_{A \times B} W = \int_{A \times B} \widetilde{W}.$$

Take any $g_A \in \mathbb{E}(I_A|\mathcal{A})$ and let

$$\begin{aligned} U_A(y) &:= \int_A W(x, y) d\mu(x) = \int W(x, y) I_A(x) d\mu(x), \\ V_A(x) &:= \int W(x, y) g_A(y) d\mu(y). \end{aligned}$$

Clearly, g_A is \mathcal{A} -measurable. Since W is $\mathcal{F} \bar{\otimes} \mathcal{A}$ -measurable (as it was noted after Claim 1), U_A is \mathcal{A} -measurable by Fubini's theorem. Similarly, V_A is also \mathcal{A} -measurable. Repeatedly using Fubini's theorem and (8), we get

$$\begin{aligned} \int_{A \times B} W(x, y) d\mu(x) d\mu(y) &= \int U_A(y) I_B(y) d\mu(y) = \\ \int U_A(y) g_B(y) d\mu(y) &= \int W(x, y) I_A(x) g_B(y) d\mu(x) d\mu(y) = \\ \int V_B(x) I_A(x) d\mu(x) &= \int V_B(x) g_A(x) d\mu(x) = \\ \int W(x, y) g_A(x) g_B(y) d\mu(x) d\mu(y) &= \int \widetilde{W}(x, y) g_A(x) g_B(y) d\mu(x) d\mu(y). \end{aligned}$$

Observe that $g_A(x) g_B(y)$ is a conditional expectation of $I_A(x) I_B(y)$ with respect to $\mathcal{A} \otimes \mathcal{A}$ while \widetilde{W} is measurable in this σ -algebra. Thus we can replace $g_A(x) g_B(y)$ by $I_A(x) I_B(y)$ in the last integral, obtaining $\int_{A \times B} \widetilde{W}$ as desired. \square

Thus all assumptions of Lemma 3.3 are satisfied and we obtain that $W \sim U^f$ for some almost kernel U on $([0, 1], \mathcal{B}, \lambda, <)$. By Lemma 3.1, we can change U on a null set so that $([0, 1], \mathcal{B}, \lambda, <, U)$ is a kernel. Clearly, the equivalence $W \sim U^f$ is not affected by this. This finishes the proof of Theorem 1.9.

6. A FINITE SZEMERÉDI-TYPE REGULARITY LEMMA FOR POSETS

In this section we prove a Szemerédi-type Regularity Lemma for posets, Theorem 6.1. (See Proemel, Steger, and Taraz [16] and Patel [15] for other versions.) We then show in Section 7 that this result can be used to answer Janson's question.

Suppose that (P, \prec) is a poset. For two disjoint sets $X, Y \subseteq P$ we write $X \not\prec Y$ if there are no $x \in X$ and $y \in Y$ such that $x \prec y$. An (ordered) partition $\mathcal{P} = (V_1, \dots, V_k)$ of the ground set P is a *poset partition* if

$$(15) \quad V_i \not\prec V_j, \quad \text{for all } 1 \leq j < i \leq k.$$

In other words, every \prec -relation that involves vertices from two different parts goes “forward”. We refer to members of \mathcal{P} as *clusters*. Let us say that \mathcal{R} is a *poset refinement* of \mathcal{P} if \mathcal{R} is a poset partition that refines \mathcal{P} (that is, for each $X \in \mathcal{R}$ there exists $Y \in \mathcal{P}$ such that $X \subseteq Y$). The *restriction* of \mathcal{P} to $X \subseteq P$ is $\mathcal{P}|_X = (V_1 \cap X, \dots, V_k \cap X)$. (For notational convenience, we allow empty parts.)

Let $G = G_{P, \prec}$ be an (undirected) graph on the vertex set P with edge set

$$E(G) := \{ \{x, y\} : x \prec y \text{ or } y \prec x \}.$$

Clearly, if we know G and a poset partition \mathcal{P} , then we can reconstruct \prec except for pairs lying inside a part. The main idea behind our Regularity Lemma is to find a poset partition of P that is regular with respect to G .

The following definitions apply to $A, B \subseteq P$. The *density* of the pair (A, B) is

$$d(A, B) := \frac{e(A, B)}{|A| |B|} := \frac{|\{(x, y) \in A \times B : x \prec y\}|}{|A| |B|}, \quad \text{if } A, B \neq \emptyset,$$

and $d(A, B) := 0$ otherwise. The pair (A, B) is called ε -*regular* if $|d(A, B) - d(X, Y)| < \varepsilon$ for each $X \subseteq A$ and $Y \subseteq B$ with $|X| \geq \varepsilon |A|$ and $|Y| \geq \varepsilon |B|$. When we will apply the definition of ε -regularity to (A, B) , it will always be the case that $B \not\prec A$ (and we obtain the standard graph definition). Also, let

$$q(A, B) := \frac{|A| |B|}{|P|^2} d^2(A, B).$$

For disjoint sets $V_1, \dots, V_k, U_1, \dots, U_m \subseteq P$, we define

$$\begin{aligned} q((V_1, \dots, V_k), (U_1, \dots, U_m)) &:= \sum_{i=1}^k \sum_{j=1}^m q(V_i, U_j), \\ q((V_1, \dots, V_k)) &:= \sum_{i < j} q(V_i, V_j). \end{aligned}$$

The function q is called the *index* and is crucial in the proof of the original Regularity Lemma. Also, let $\mathbb{I}_\varepsilon((V_1, \dots, V_k))$ be the set of pairs (i, j) such that $i < j$ and (V_i, V_j) is not ε -regular.

The sizes of the clusters in our Regularity Lemma can vary vastly (at least in our proof). This is why our next definition is slightly different from the standard one. A poset partition $\mathcal{P} = (V_1, \dots, V_k)$ of P is ε -regular if each $|V_i| \leq \max(\varepsilon|P|, 1)$ and

$$\sum_{(i,j) \in \mathbb{I}_\varepsilon(\mathcal{P})} |V_i| |V_j| \leq \varepsilon \binom{|P|}{2}.$$

Theorem 6.1 (Regularity Lemma for Posets). *For each $\varepsilon > 0$ there exists a number M such that the following holds. For each poset (P, \prec) with a poset partition \mathcal{P} such that $|\mathcal{P}| \leq 1/\varepsilon$, there exists a poset refinement \mathcal{R} of \mathcal{P} which is ε -regular and has at most M parts.*

Remark 6.2. It is important for our later application in Section 7 that there is no garbage cluster in our partition.

We prove Theorem 6.1 by following Szemerédi's original proof of the Regularity Lemma for graphs [20] (a more accessible reference is for example [7, Section 7.4]). The basic idea is that if a current partition \mathcal{P} is not ε -regular then we can refine it so that $q(\mathcal{P})$ increases by at least δ , where $\delta > 0$ depends on ε only. Since q is always between 0 and 1, we reach an ε -regular partition in at most $1/\delta$ refinements. The following “pumping-up” lemma estimates by how much we can increase q by subdividing one irregular pair (A, B) .

Lemma 6.3. *Suppose that (P, \prec) is a poset and $A, B \subseteq P$ are disjoint nonempty sets. If $B \not\prec A$ and (A, B) is not ε -regular, then there are partitions $A = Z_1 \cup Z_2$ and $B = Z_3 \cup Z_4$ such that $Z_2 \not\prec Z_1$, $Z_4 \not\prec Z_3$, and*

$$(16) \quad q((Z_1, Z_2), (Z_3, Z_4)) \geq q(A, B) + \varepsilon^4 \frac{|A| |B|}{n^2}.$$

Proof. Let $d := d(A, B)$. Consider a witness of irregularity (X, Y) of the pair (A, B) . Assume without loss of generality that $d(X, Y) \geq d + \varepsilon$.

Iteratively, repeat the following as long as possible: replace some $x \in X$ by some $y \in A \setminus X$ with $y \prec x$. Clearly, this operation preserves the size of X and cannot decrease $d(X, Y)$. Also, we have to stop at some point. Let Z_1 be the final X and let $Z_2 := A \setminus Z_1$.

Similarly, replace $Y \subseteq B$ by an up-closed subset $Z_4 \subseteq B$ such that $|Z_4| = |Y|$ and $d(Z_1, Z_4) \geq d(Z_1, Y) \geq d + \varepsilon$. Let $Z_3 := B \setminus Z_4$.

Of course, we have that $Z_2 \not\prec Z_1$ and $Z_4 \not\prec Z_3$. Also, (Z_1, Z_4) witnesses that the pair (A, B) is not ε -regular. Now, (16) follows from the proof of Lemma 7.4.3 in [7]. \square

Proof of Theorem 6.1. Let $s := \lceil 4/\varepsilon^5 \rceil$, $k_0 := \lceil 2/\varepsilon \rceil$, and inductively for $t = 0, \dots, s-1$, let $k_{t+1} := k_t 2^{k_t-1}$. We claim that $M = k_s$ suffices.

Suppose that $n > 1/\varepsilon$ for otherwise we can let \mathcal{R} be a partition into singletons.

Initially, let \mathcal{R}_0 be an arbitrary poset refinement of \mathcal{P} such that $|\mathcal{R}_0| \leq k_0$ and each part has at most εn vertices.

Iteratively, for $t = 0, 1, \dots$, we repeat the following procedure. Let $\mathcal{R}_t = (V_1, \dots, V_k)$. If \mathcal{R}_t is ε -regular then we stop and output \mathcal{R}_t ; so suppose otherwise. Let $\mathcal{R}' := \mathcal{R}_t$. We modify \mathcal{R}' by using another (embedded) iterative procedure. Namely, in turn for each $(i, j) \in \mathbb{I}_\varepsilon(\mathcal{R}_t)$, we take the partitions $V_i = Z_{1ij} \cup Z_{2ij}$ and $V_j = Z_{3ij} \cup Z_{4ij}$ returned by Lemma 6.3 and replace every $X \in \mathcal{R}'$ by $X \cap Z_{1ij}, \dots, X \cap Z_{4ij}$, with these four parts coming in the specified order. Clearly, \mathcal{R}' is still a poset partition. Once we have processed all elements of $\mathbb{I}_\varepsilon(\mathcal{R}_t)$, we let $\mathcal{R}_{t+1} := \mathcal{R}'$.

In order to estimate how q changes, let us re-write

$$(17) \quad q(\mathcal{R}_{t+1}) - q(\mathcal{R}_t) = \sum_{1 \leq i < j \leq k} \left(q(\mathcal{R}_{t+1}|_{V_i}, \mathcal{R}_{t+1}|_{V_j}) - q(V_i, V_j) \right).$$

We can estimate each summand corresponding to $(i, j) \in \mathbb{I}_\varepsilon(\mathcal{R}_t)$ by passing from $q(V_i, V_j)$ first to $q((Z_{1ij}, Z_{2ij}), (Z_{3ij}, Z_{4ij}))$ and then to $q(\mathcal{R}_{t+1}|_{V_i}, \mathcal{R}_{t+1}|_{V_j})$. The first step increases q as specified by Lemma 6.3. The second step has non-negative effect by [7, Lemma 7.4.2]. Each other term in the right-hand side of (17) is non-negative, again by [7, Lemma 7.4.2]. Since \mathcal{R}_t is not ε -regular, we conclude that

$$(18) \quad q(\mathcal{R}_{t+1}) - q(\mathcal{R}_t) \geq \frac{\varepsilon^4}{n^2} \sum_{(i,j) \in \mathbb{I}_\varepsilon(\mathcal{R}_t)} |V_i| |V_j| > \frac{\varepsilon^4}{n^2} \varepsilon \binom{n}{2} \geq \frac{\varepsilon^5}{4}.$$

We have that $q(\mathcal{P}) \leq 1$ for any partition \mathcal{P} (while trivially $q(\mathcal{P}) \geq 0$), see e.g. [7, Page 174]. By (18), we repeat the iteration procedure at most s times before we reach an ε -regular poset partition. As each part of \mathcal{R}^t is split into at most $2^{|\mathcal{R}_t|-1}$ parts, we have that $|\mathcal{R}^{t+1}| \leq |\mathcal{R}_t| 2^{|\mathcal{R}_t|-1}$. Thus the final partition has at most M parts, as required. \square

If we do not know \prec but know an ε -regular partition $\mathcal{R} = (V_1, \dots, V_k)$ and the densities between all pairs of parts, then we can still derive various information about the poset P . For example, given two subsets $S, T \subseteq P$, one would expect to see approximately

$$e'(S, T) := \sum_{i < j} d(V_i, V_j) |V_i \cap S| |V_j \cap T|$$

directed arcs from S to T . Indeed, this is the case for posets.

Lemma 6.4. *Given the above assumptions, we have*

$$(19) \quad |e(S, T) - e'(S, T)| \leq 3\varepsilon \binom{|P|}{2}.$$

Proof. Let $n := |P|$. Assuming the worst-case scenario, the edges inside a part or inside a non- ε -regular pair contribute at most $\varepsilon \binom{n}{2} + \varepsilon \binom{n}{2}$ to the left-hand side of (19). For every ε -regular pair (V_i, V_j) with $i < j$, we have

$$\left| e(V_i \cap S, V_j \cap T) - d(V_i, V_j) |V_i \cap S| |V_j \cap T| \right| \leq \varepsilon |V_i| |V_j|.$$

Indeed, if $|S \cap V_i| |T \cap V_j| \leq \varepsilon |V_i| |V_j|$, then we are trivially done; otherwise both S and T take more than ε -proportion of respectively V_i and V_j and the bound follows by the ε -regularity of (V_i, V_j) . Thus the aggregate contribution of ε -regular pairs to (19) is at most $\varepsilon \binom{n}{2}$. \square

7. AN ALTERNATIVE PROOF OF THEOREM 1.7

Let $\{(P_n, \prec_n)\}_{n \in \mathbb{N}}$ be a convergent sequence of posets. We have to construct a kernel $([0, 1], \mathcal{B}, \lambda, <, U)$ such that for every poset P we have

$$(20) \quad t(P, U) = \lim_{n \rightarrow \infty} t(P, P_n).$$

We construct U following closely the analogous construction of Lovász and Szegedy [13, Theorem 2.4] (see also [14, Theorem 5.1]). In brief, the proof proceeds by finding a $\frac{1}{k}$ -regular partition $\mathcal{P}_{n,k}$ of P_n with the number of parts bounded by a function of k only. Then we construct a step-function $W_{n,k} : [0, 1]^2 \rightarrow [0, 1]$ that encodes the part ratios and densities of $\mathcal{P}_{n,k}$. Since the ‘‘complexity’’ of $W_{n,k}$ is bounded by a function of k , a diagonalisation process gives a subsequence $\{P_{n_i}\}_{i \in \mathbb{N}}$ such that, for every k , we have $W_{n,k} \rightarrow U_k$ a.e. for some $U_k : [0, 1]^2 \rightarrow [0, 1]$. Additionally, when we choose our partitions $\mathcal{P}_{k,n}$, we can assume that they are nested for each n . This allows us to write U_{k-1} as a conditional expectation of U_k and conclude that $\{U_k\}_{k \in \mathbb{N}}$ converges to some U a.e. Finally, we need to apply Lemma 3.1 to transform an almost kernel U into a kernel.

Let us give more details. Let $m_1 = 1$ and inductively for $k = 2, 3, \dots$ let m_k be sufficiently large such that every poset partition with at most m_{k-1} parts admits a $\frac{1}{k}$ -regular poset refinement with m_k parts. Such a number exists by Theorem 6.1. (Recall that we allow empty parts.) For each $n \in \mathbb{N}$, let $\mathcal{P}_{n,1} := (P_n)$ be the trivial partition and then inductively for $k = 2, 3, \dots$ let

$$(21) \quad \mathcal{P}_{n,k} = (V_{n,k,1}, \dots, V_{n,k,m_k})$$

be a $\frac{1}{k}$ -regular poset partition of (P_n, \prec_n) that refines $\mathcal{P}_{n,k-1}$. This nestedness allows us for each n , to choose a total ordering \prec'_n of (P_n, \prec_n) which is *compatible* with every poset partition $\mathcal{P}_{n,k}$ (that is, $V_{n,k,i} \not\prec'_n V_{n,k,j}$ whenever $i > j$). By relabelling, let us assume that $P_n = \{1, \dots, |P_n|\}$ and \prec'_n is the standard order.

Already at this point, it makes sense to start operating with functions. Let $W_n : [0, 1]^2 \rightarrow \{0, 1\}$ be the step-function that encodes the \prec_n -relation in the obvious way: W_n is constant on $[\frac{i-1}{v}, \frac{i}{v}) \times [\frac{j-1}{v}, \frac{j}{v})$, where $v := |P_n|$, and assumes value 1 there if and only if $i \prec_n j$. It is easy to see that

$$t(P, P_n) = t(P, W_n), \quad \text{for every poset } P$$

where we view W_n as a kernel on $([0, 1], \mathcal{B}, \lambda, <)$.

Let $\mathcal{P}'_{n,k} = (V'_{n,k,1}, \dots, V'_{n,k,m_k})$ be the partition of $[0, 1]$ into consecutive intervals corresponding to (21). (Thus, for example, $\lambda(V'_{n,k,i}) = |V_{n,k,i}|/|P_n|$.) Let $W_{n,k}$ be the step-function on $\mathcal{P}'_{n,k} \times \mathcal{P}'_{n,k}$, whose steps correspond to the parts of $\mathcal{P}_{n,k}$ and whose values correspond to densities between parts. We can write this more compactly as

$$W_{n,k} \sim \mathbb{E}(W_n | \sigma(\mathcal{P}'_{n,k})),$$

a conditional expectation of W_n with respect to the (finite) σ -algebra generated by $\mathcal{P}'_{n,k}$. Since $\sigma(\mathcal{P}'_{n,1}) \subseteq \sigma(\mathcal{P}'_{n,2}) \subseteq \dots$, we have

$$W_{n,k} \sim \mathbb{E}(W_{n,k+1} | \sigma(\mathcal{P}'_{n,k})), \quad k \geq 1,$$

which translates into the combinatorially obvious fact that the densities of $\mathcal{P}_{n,k}$ can be obtained by averaging over the densities in the finer partition $\mathcal{P}_{n,k+1}$.

Since each $W_{n,k}$ can be described by specifying part sizes and densities (which involves at most $m_k + \binom{m_k}{2}$ reals in $[0, 1]$), the standard diagonalisation process gives a subsequence $\{n_i\}_{i \in \mathbb{N}}$ such that these parameters converge for every k . Thus $W_{n_i,k} \rightarrow U_k$ a.e. for some step-function U_k with m_k steps that are intervals and are ordered as $\mathcal{P}'_k = (V'_{k,1}, \dots, V'_{k,m_k})$. Since $\{P_n\}_{n \in \mathbb{N}}$ is convergent, passing to a subsequence does not affect (20); thus we can assume that $\{W_{n,k}\}_{n \in \mathbb{N}}$ itself a.e. converges to U_k . Clearly, $\sigma(\mathcal{P}'_1) \subseteq \sigma(\mathcal{P}'_2) \subseteq \dots$ a.e. and

$$U_k \sim \mathbb{E}(U_{k+1} | \sigma(\mathcal{P}'_k)).$$

Thus, by the Martingale Convergence Theorem (see e.g. [2, Theorem 10.3.3]), $U_k \rightarrow U$ a.e. for some $U : [0, 1]^2 \rightarrow [0, 1]$.

The obtained function U , as the a.e. pointwise limit of Borel functions, is Borel a.e. Clearly, the kernel axioms hold for $([0, 1], \mathcal{B}, \lambda, U)$ for all inputs that do not require the evaluation of U on a point of

$$X := \{(x, y) \in [0, 1]^2 : U_k(x, y) \not\rightarrow U(x, y) \text{ or } \exists k \ W_{n,k}(x, y) \not\rightarrow U_k(x, y)\},$$

the set where some convergence fails. Since X has measure zero, U is an almost kernel. By applying Lemma 3.1, we can assume that U is a kernel.

It remains to show that (20) holds. The *cut-norm* of a bounded measurable function $W : [0, 1]^2 \rightarrow \mathbb{R}$ is defined by

$$(22) \quad \|W\|_{\square} = \sup_{S, T \in \mathcal{B}} \left| \int_{S \times T} W(x, y) d\lambda(x) d\lambda(y) \right|.$$

Claim 1. $\|W_n - W_{n,k}\|_{\square} \leq \frac{5}{2k}$ for any $k, n \in \mathbb{N}$.

Proof of Claim. Let $W := W_n - W_{n,k}$. Assume that $v := |P_n| > k$ for otherwise there is nothing to do as $W = 0$.

Observe that, up to an additive error $\frac{1}{v}$, it is enough to consider those S and T in (22) that are unions of intervals $V_i := [\frac{i-1}{v}, \frac{i}{v})$ for $i \in [v]$. Indeed, fix any $S, T \in \mathcal{B}$ with, say, $\int_{S \times T} W \geq 0$ and take $i \in [v]$ one by one. If we modify S and T inside V_i , then the integral of W over

$$((V_i \times V_i^c) \cup (V_i^c \times V_i)) \cap (S \times T)$$

is a linear function of $\lambda(V_i \cap S)$ and $\lambda(V_i \cap T)$. Thus we can make each of these to belong to $\{0, 1/v\}$ without decreasing the above contribution. Updating S and T accordingly, we decrease $\int_{S \times T} W$ by at most $\int_{V_i \times V_i} |W| \leq 1/v^2$. When we have iteratively processed all $i \in [v]$, both S and T have the desired form.

Thus, by (19), we obtain the required:

$$\|W\|_{\square} \leq \frac{3}{k} \binom{v}{2} \frac{1}{v^2} + \frac{1}{v} \leq \frac{3}{2k} + \frac{1}{k} = \frac{5}{2k}.$$

□

Now, we are ready to verify (20). Take any poset (P, \prec) and $\varepsilon > 0$. Let $m := e(G_{P, \prec})$ be the number of pairs in \prec .

Since we deal with bounded measurable functions, all convergences also hold in the ℓ_1 -space on $([0, 1]^2, \mathcal{B}, \lambda)$ by [2, Theorem 2.2.3]. Thus there is $k \geq \frac{15m}{2\varepsilon}$ such that $\|U - U_k\|_1 \leq \frac{\varepsilon}{3m}$ and, fixing this k , there is n_0 such that $\|U_k - W_{n,k}\|_1 \leq \frac{\varepsilon}{3m}$ for all $n \geq n_0$. Clearly, $\|f\|_{\square} \leq \|f\|_1$ for any integrable f . Thus, by the Triangle Inequality and Claim 1, we have that for all $n \geq n_0$

$$\begin{aligned} \|U - W_n\|_{\square} &\leq \|U - U_k\|_{\square} + \|U_k - W_{n,k}\|_{\square} + \|W_{n,k} - W_n\|_{\square} \\ &\leq \|U - U_k\|_1 + \|U_k - W_{n,k}\|_1 + \frac{5}{2k} \leq \frac{\varepsilon}{m}. \end{aligned}$$

By [11, Lemma 6.4], we have that $|t(P, U) - t(P, W_n)| \leq m \|U - W_n\|_{\square} \leq \varepsilon$. Since ε and P were arbitrary, (20) follows.

Summarising, $([0, 1], \mathcal{B}, \lambda, <, U)$ is a kernel that establishes Theorem 1.7.

Remark 7.1. An alternative way to proving that the densities of F in W_n and $W_{n,k}$ are close is to adopt the Counting Lemma (see e.g. [18, Theorem 5]) to our settings. We do not see any principal difficulties here but we expect that the error term would be larger.

Remark 7.2. In the above proof it is not generally true that $\|W_n - W_{n,k}\|_1$ is small for sufficiently large k : for example, $W_{n,k}$ may be strictly between 0 and 1 on a set of positive measure (while W_n is always $\{0, 1\}$ -valued).

8. CONCLUDING REMARKS

There are two natural ways to extend the definition of convergence to the case when the poset orders do not tend to infinity. One is to just use (1). Another, adopted by Janson [11, Definition 3.2], is to say that $\{P_n\}_{n \in \mathbb{N}}$ with $|P_n| \not\rightarrow \infty$ is convergent if the sequence is eventually constant (up to isomorphism). The choice of which one to use (or none) is more a matter of convenience. For example, this choice may depend on whether we want the “limits” of (P, P, \dots) and $(P^{(1)}, P^{(2)}, \dots)$ to be the same or not. Here the *blow-up* $P^{(k)}$ of P is obtained by cloning each vertex of P k times; obviously, $t(Q, P) = t(Q, P^{(k)})$ for every poset Q . Since all results stated in the Introduction can be trivially reduced to the case $|P_n| \rightarrow \infty$ by blowing posets up, we decided to use Definition 1.1.

Of course, the assumption that (S, \mathcal{F}, μ) is atomless is necessary in Theorems 1.9 and 1.10. This assumption can be removed if we are allowed to modify $([0, 1], \mathcal{B}, \lambda, <)$ by shifting positive measure to a some countable subset $X \subseteq [0, 1]$, where X depends on $(S, \mathcal{F}, \mu, \triangleleft)$. However, we believe that the versions presented in the Introduction are neater.

We cannot require in Theorems 1.9 and 1.10 that f preserves *every* relation (i.e. that the set in (7) is empty) as the following example demonstrates. Let $S := [0, 1]$ with the Lebesgue measure λ on the Borel σ -algebra \mathcal{B} . Fix an irrational number τ . Let $T : S \rightarrow S$ map x to $x + \tau \pmod{1}$. If we view S as a circle, then T is an aperiodic rotation. Define $x \triangleleft y$ if there is $k \in \mathbb{N}$ with $y = T^k(x)$. The constructed relation \triangleleft is a Borel subset of S^2 (of measure zero). Let us suppose on the contrary that there is an inclusion $f : (S, \mathcal{B}, \lambda, \triangleleft) \rightarrow ([0, 1], \mathcal{B}, \lambda, <)$ such that the set in (7) is empty. Let $A := f^{-1}([0, \frac{1}{2}])$. Since A is a down-closed set with respect to $<$, we have that $T^{-1}(A) \subseteq A$. Since T is measure preserving, we conclude that $T^{-1}(A) \sim A$. However, this contradicts the well-known fact (see e.g. [2, Example 10.9.9]) that T is ergodic. Alternatively, let $B := \bigcap_{k=1}^{\infty} T^{-k}(A)$. Then B is a measurable set such that $T^{-1}(B) = B$ (exactly) and $\mu(B) = 1/2$ (by σ -additivity). The same applies to B^c . By taking density points x and y of B and B^c respectively and a sequence of k such that $T^k(x) \rightarrow y$, one readily arrives at the desired contradiction.

Also, the assumption that W is strict in Theorem 1.9 is needed. For example, take $[0, 1]^2$ with the Legesgue measure on the Borel sets and let $(x, y) \triangleleft (x', y')$ if $x < x'$. Let, for example, $W((x, y), (x', y'))$ be y' if $x' > x + 1/2$ and 0 otherwise. It is easy to see that every inclusion of this ordered probability space into the unit interval is a.e. equal to the projection onto the first coordinate. However, $W((x, y), (x', y'))$ is essentially non-constant on (x, x') -slices for $x' > 1/2 + x$ and thus cannot be equivalent to some pull-back.

REFERENCES

1. T. Austin, *On exchangeable random variables and the statistics of large graphs and hypergraphs*, Probab. Surv. **5** (2008), 80–145.
2. V. I. Bogachev, *Measure theory*, Springer, Berlin, 2007.

3. C. Borgs, J. Chayes, and L. Lovász, *Moments of two-variable functions and the uniqueness of graph limits*, Geom. Func. Analysis **19** (2010), 1597–1619.
4. C. Borgs, J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, *Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing*, Adv. Math. **219** (2008), 1801–1851.
5. G. Brightwell and N. Georgiou, *Continuum limits for classical sequential growth models*, Random Struct. Algorithms **36** (2010), 218–250.
6. P. Diaconis and S. Janson, *Graph limits and exchangeable random graphs*, Rendiconti di Matematica **28** (2008), 33–61.
7. R. Diestel, *Graph theory*, 4th ed., Springer, Berlin, 2010.
8. G. Elek and B. Szegedy, *A measure-theoretic approach to the theory of dense hypergraphs*, Adv. Math. **231** (2012), 1731–1772.
9. C. Hoppen, Y. Kohayakawa, C. G. Moreira, B. Ráth, and R. M. Sampaio, *Limits of permutation sequences*, E-Print [arXiv.org:1103.5844](https://arxiv.org/abs/1103.5844), 2011.
10. C. Hoppen, Y. Kohayakawa, C. G. Moreira, and R. M. Sampaio, *Limits of permutation sequences through permutation regularity*, E-Print [arXiv.org:1106.1663](https://arxiv.org/abs/1106.1663), 2011.
11. S. Janson, *Posets limits and exchangeable random posets*, Combinatorica **31** (2011), 529–563.
12. ———, *Limits of interval orders and semiorders*, J. Combin. **3** (2012), 163–184.
13. L. Lovász and B. Szegedy, *Limits of dense graph sequences*, J. Combin. Theory (B) **96** (2006), 933–957.
14. ———, *Szemerédi’s Lemma for the analyst*, Geom. Func. Analysis **17** (2007), 252–270.
15. V. Patel, *Partitions of combinatorial structures*, PhD Thesis, London School of Economics, 2009.
16. H. J. Prömel, A. Steger, and A. Taraz, *Counting partial orders with a fixed number of comparable pairs*, Combin. Probab. Computing **10** (2001), 159–177.
17. A. Razborov, *Flag algebras*, J. Symb. Logic **72** (2007), 1239–1282.
18. M. Simonovits and V. T. Sós, *Szemerédi’s partition and quasirandomness*, Random Struct. Algorithms **2** (1991), 1–10.
19. B. Szegedy, *Gowers norms, regularization and limits of functions on abelian groups*, E-Print [arxiv.org:1010.6211](https://arxiv.org/abs/1010.6211), 2010.
20. E. Szemerédi, *Regular partitions of graphs*, Proc. Colloq. Int. CNRS, Paris, 1976, pp. 309–401.
21. T. Tao, *A correspondence principle between (hyper)graph theory and probability theory, and the (hyper)graph removal lemma*, J. Anal. Math. **103** (2007), 1–45.

MATHEMATICS INSTITUTE, UNIVERSITY OF WARWICK, COVENTRY, CV4 7AL, UNITED KINGDOM

E-mail address: J.Hladky@warwick.ac.uk

MATHEMATICS INSTITUTE, UNIVERSITY OF WARWICK, COVENTRY, CV4 7AL, UNITED KINGDOM

E-mail address: A.Mathe@warwick.ac.uk

SCHOOL OF MATHEMATICS, BIRMINGHAM UNIVERSITY, EDGBASTON, BIRMINGHAM B15 2TT, UNITED KINGDOM

E-mail address: viresh.s.patel@googlemail.com

MATHEMATICS INSTITUTE, UNIVERSITY OF WARWICK, COVENTRY, CV4 7AL, UNITED KINGDOM

URL: <http://homepages.warwick.ac.uk/staff/O.Pikhurko/>